Bayesian Analysis for Complex Physical Systems Modeled by Computer Simulators: Current Status and Future Challenges

Michael Goldstein
Department of Mathematical Sciences,
Durham University *

* Thanks to EPSRC, NERC for funding, and to Jonathan Cumming, Ian Vernon, Danny Williamson

Complex physical models

Most large and complex physical systems are studied by mathematical models, implemented as high dimensional computer simulators (like climate models).

To use complex simulators to make statements about physical systems (like climate), we need to quantify the uncertainty involved in moving from the model to the system.

These questions are both

practical/methodological (how can we work out what climate is likely to be?)

and

foundational (why should our methods work and what do our answers mean?)

Examples

Oil reservoirs An oil reservoir simulator is used to manage assets associated with the reservoir.
The aim is commercial: to develop efficient production schedules, determine whether and where to sink new wells, and so forth.

Galaxy formation The study of the development of the Universe is carried out by using a Galaxy formation simulator.
The aim is scientific - to gain information about the physical processes underlying the Universe.

Climate change Large scale climate simulators are constructed to assess likely effects of human intervention upon future climate behaviour.
Aims are both scientific - much is unknown about the large scale interactions which determine climate - and also very practical, as such simulators provide evidence for the importance of changing human behaviour before possibly irreversible changes are set into motion.

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)

(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),

(iii) functional uncertainty (model evaluations take a long time, so the function is unknown almost everywhere)

(iv) stochastic uncertainty (either the model is stochastic, or it should be),

(v) solution uncertainty (as the system equations can only be solved to some necessary level of approximation).

(vi) structural uncertainty (the model only approximates the physical system),

(vii) measurement uncertainty (as the model is calibrated against system data all of which is measured with error),

(viii) multi-model uncertainty (usually we have not one but many models related to the physical system)

(ix) decision uncertainty (to use the model to influence real world outcomes, we need to relate things in the world that we can influence to inputs to the simulator and through outputs to actual impacts. These links are uncertain.)

General form

Different physical models vary in many aspects, but the formal structures for analysing the physical system through computer simulators are very similar (which is why there is a common underlying methodology).

Each simulator can be conceived as a function $f(x)$, where
x: input vector, representing unknown properties of the physical system;
$f(x)$: output vector representing system behaviour.

Interest in general qualitative insights plus some of the following.
the “appropriate” (in some sense) choice, x^*, for the system properties x,
how informative $f(x^*)$ is for actual system behaviour, y.
the use that we can make of historical observations z, observed with error on a subset y_h of y, both to test and to constrain the model,
the optimal assignment of any decision inputs, d, in the model.

[In a climate model, y_h might correspond to historical climate outcomes over space and time, y to current and future climate, and the “decisions” might correspond to different policy relevant choices such as carbon emission scenarios.]

“Solving” these problems

Durham University

If observations, z, are made without error and the model is perfect reproduction of the system, we can write $z = f_h(x^*)$, invert f_h to find x^*, learn about all future components of $y = f(x^*)$ and choose decision elements of x^* to optimise properties of y.

COMMENT: This would be very hard.

In practice, the observations z are made with error, and model is not the same as physical system so we must separate the uncertainty representation into two relations and carry out statistical inversion/optimisation:

$$z = y_h \oplus e, \;\; y = f(x^*) \oplus \epsilon$$

where e, ϵ have some appropriate probabilistic specification, possibly involving parameters which require estimation.

COMMENT: This is much harder.

COMMENT And we still haven’t accounted for condition uncertainty, multi-model uncertainty, etc.

Current state of the art

Many people work on different aspects of these uncertainty analyses

Great resource: the Managing Uncertainty in Complex Models web-site

http://www.mucm.ac.uk/ (for references, papers, toolkit, etc.)

[MUCM is a consortium of U. of Aston, Durham, LSE, Sheffield, Southampton - with Basic Technology funding.]

However, in practice, it is extremely rare to find a serious quantitification of the total uncertainty about a complex system arising from the all of the uncertainties in the model analysis.

Therefore, for all applications, no-one really knows the reliability of the model based analysis. Therefore, there is no sound basis for identifying appropriate real world decisions based on such model analyses.

This is because

modellers/scientists don't think about total uncertainty this way

nor do most statisticians

policy makers don't know how to frame the right questions for the modellers

there are few funding mechanisms to support this activity

and it is hard!

RAPID-WATCH

What are the implications of RAPID-WATCH observing system data and other recent observations for estimates of the risk due to rapid change in the MOC? In this context risk is taken to mean the probability of rapid change in the MOC and the consequent impact on climate (affecting temperatures, precipitation, sea level, for example). This project must:

* contribute to the MOC observing system assessment in 2011;
* investigate how observations of the MOC can be used to constrain estimates of the probability of rapid MOC change, including magnitude and rate of change;
* make sound statistical inferences about the real climate system from model simulations and observations;
* investigate the dependence of model uncertainty on such factors as changes of resolution;
* assess model uncertainty in climate impacts and characterise impacts that have received less attention (eg frequency of extremes).

The project must also demonstrate close partnership with the Hadley Centre.

RAPID-WATCH

What are the implications of RAPID-WATCH observing system data and other recent observations for estimates of the risk due to rapid change in the MOC? In this context risk is taken to mean the probability of rapid change in the MOC and the consequent impact on climate (affecting temperatures, precipitation, sea level, for example). This project must:

* contribute to the MOC observing system assessment in 2011;
* investigate how observations of the MOC can be used to constrain estimates of the probability of rapid MOC change, including magnitude and rate of change;
* make sound statistical inferences about the real climate system from model simulations and observations;
* investigate the dependence of model uncertainty on such factors as changes of resolution;
* assess model uncertainty in climate impacts and characterise impacts that have received less attention (eg frequency of extremes).

The project must also demonstrate close partnership with the Hadley Centre.

RAPID-WATCH

What are the implications of RAPID-WATCH observing system data and other recent observations for estimates of the risk due to rapid change in the MOC? In this context risk is taken to mean the probability of rapid change in the MOC and the consequent impact on climate (affecting temperatures, precipitation, sea level, for example). This project must:

* contribute to the MOC observing system assessment in 2011;
* investigate how observations of the MOC can be used to constrain estimates of the probability of rapid MOC change, including magnitude and rate of change;
* make sound statistical inferences about the real climate system from model simulations and observations;
* investigate the dependence of model uncertainty on such factors as changes of resolution;
* assess model uncertainty in climate impacts and characterise impacts that have received less attention (eg frequency of extremes).

The project must also demonstrate close partnership with the Hadley Centre.

Subjectivist Bayes

Durham University

In the subjectivist Bayes view, the meaning of any probability statement is the uncertainty judgement of a specified individual, expressed on the scale of probability (by consideration of some operational elicitation scheme, for example by consideration of betting preferences).

This interpretation has an agreed testable meaning, sufficiently precise to act as the basis of a discussion about the meaning of the analysis.

In this interpretation, any probability statement is the judgement of a named individual, so we should speak not of the probability of rapid climate change, but instead of Anne's probability or Bob's probability of rapid climate change and so forth.

There is a big issue of perception here, as most people expect something more authoritative and objective than a probability which is one person's judgement.

However, the disappointing thing is that, in almost all cases, stated probabilities emerging from a complex analysis are not even the judgements of any individual.

So, it is not unreasonable that the objective of our analysis should be probabilities which are asserted by at least one person (more would be good!).

Bayesian uncertainty analysis for complex models

Durham University

Aim: to tackle previously intractable problems arising from the uncertainties inherent in imperfect computer models of highly complex physical systems, using a Bayesian formulation. This involves

- prior probability distribution for best inputs x^*
- a probabilistic uncertainty description for the computer function f
- a probabilistic discrepancy measure relating $f(x^*)$ to the system y
- a likelihood function relating historical data z to y

This full probabilistic description provides a formal framework to synthesise expert elicitation, historical data and a careful choice of simulator runs.
We may then use our collection of computer evaluations and historical observations to analyse the physical process to

- determine values for simulator inputs (calibration; history matching);
- assess the future behaviour of the system (forecasting).
- “optimise” the performance of the system

Approaches for Bayesian analysis

Within the Bayesian approach, we have two choices.

(i) Full Bayes analysis, with complete joint probabilistic specification of all of the uncertain quantities in the problem

or

(ii) Bayes linear analysis, based on a prior specification of the means, variances and covariances of all quantities of interest, where we make expectation, rather than probability, the primitive for the theory, following de Finetti “Theory of Probability”(1974,1975).

de Finetti chooses expectation over probability as, if expectation is primitive, then we can choose to make as many or as few expectation statements as we choose, whereas, if probability is primitive, then we must make all of the probability statements before we can make any of the expectation statements, so that we have the option of restricting our attention to whatever subcollection of specifications we are interested in analysing carefully.

Full Bayes analysis can be more informative if done extremely carefully, both in terms of the prior specification and the analysis. Bayes linear analysis is partial but easier, faster, more robust particularly for history matching and forecasting.

Bayes linear approach

For very large scale problems a full Bayes analysis is very hard because
(i) it is difficult to give a meaningful full prior probability specification over high dimensional spaces;
(ii) the computations, for learning from data (observations and computer runs), particularly when choosing informative runs, may be technically difficult;
(iii) the likelihood surface is extremely complicated, and any full Bayes calculation may be extremely non-robust.
However, the idea of the Bayesian approach, namely capturing our expert prior judgements in stochastic form and modifying them by appropriate rules given observations, is conceptually appropriate (and there is no obvious alternative).
The Bayes Linear approach is (relatively) simple in terms of belief specification and analysis, as it is based only on the mean, variance and covariance specification which, following de Finetti, we take as primitive.
For a full account, see
Michael Goldstein and David Wooff (2007) Bayes Linear Statistics: Theory and Methods, Wiley.

Bayes linear adjustment

Bayes Linear adjustment of the mean and the variance of y given z is

$$\mathsf{E}_z[y] = \mathrm{E}(y) + \mathrm{Cov}(y,z)\mathrm{Var}(z)^{-1}(z - \mathrm{E}(z)),$$
$$\mathsf{Var}_z[y] = \mathrm{Var}(y) - \mathrm{Cov}(y,z)\mathrm{Var}(z)^{-1}\mathrm{Cov}(z,y)$$

$\mathsf{E}_z[y]$, $\mathsf{Var}_z[y]$ are the expectation and variance for y adjusted by z.

Bayes linear adjustment may be viewed as:

an approximation to a full Bayes analysis;

or

the “appropriate” analysis given a partial specification based on expectation as primitive (with methodology for modelling, interpretation and diagnostics).

The foundation for the approach is an explicit treatment of temporal uncertainty, and the underpinning mathematical structure is the inner product space (not probability space, which is just a special case).

Function emulation

Uncertainty analysis, for high dimensional problems, is even more challenging if the function $f(x)$ is expensive, in time and computational resources, to evaluate for any choice of x. [For example, large climate models.]

In such cases, f must be treated as uncertain for all input choices except the small subset for which an actual evaluation has been made.

Therefore, we must construct a description of the uncertainty about the value of $f(x)$ for each x.

Such a representation is often termed an emulator of the function - the emulator both suggests an approximation to the function and also contains an assessment of the likely magnitude of the error of the approximation.

We use the emulator either to provide a full joint probabilistic description of all of the function values (full Bayes) or to assess expectations variances and covariances for pairs of function values (Bayes linear).

Form of the emulator

We may represent beliefs about component f_i of f, using an emulator:

$$f_i(x) = \sum_j \beta_{ij} g_{ij}(x) \oplus u_i(x)$$

where $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x, $u_i(x)$ is a weakly second order stationary stochastic process, with (for example) correlation function

$$\mathrm{Corr}(u_i(x), u_i(x')) = \exp(-(\tfrac{\|x-x'\|}{\theta_i})^2)$$

$Bg(x)$ expresses global variation in f. $u(x)$ expresses local variation in f

We fit the emulators, given a collection of carefully chosen model evaluations, using our favourite statistical tools - generalised least squares, maximum likelihood, Bayes - with a generous helping of expert judgement.

We need careful (multi-output) experimental design to choose informative model evaluations, and detailed diagnostics to check emulator validity.

Linked emulators

If the simulator is really slow to evaluate, then we emulate by jointly modelling the simulator with a fast approximate version, f', plus older generations of the simulator which we've already emulated and so forth.

So, for example, based on many fast simulator evaluations, we build emulator

$$f_i'(x) = \sum_j \beta_{ij}' g_{ij}(x) \oplus u_i'(x)$$

We use this form as the prior for the emulator for $f_i(x)$.

Then a relatively small number of evaluations of $f_i(x)$, using relations such as

$$\beta_{ij} = \alpha_i \beta_{ij}' + \gamma_{ij}$$

lets us adjust the prior emulator to an appropriate posterior emulator for $f_i(x)$.

[This approach exploits the heuristic that we need many more function evaluations to identify the qualitative form of the model (i.e. choose appropriate forms $g_{ij}(x)$, etc) than to assess the quantitative form of all of the terms in the model - particularly if we fit meaningful regression components.]

Illustration from RAPID (thanks to Danny Williamson)

Durham University

One of the main aims of the RAPIT programme is to assess the risk of shutdown of the AMOC (Atlantic Meridionnal Overturning Circulation)which transports heat from the tropics to Northern Europe and how this risk depends on the future emissions scenario for CO2.

RAPIT aims to use large ensembles of the UK Met Office climate model HadCM3, run through climateprediction.net

[Our first ensemble of 20,000 runs is out now.]

As a preliminary demonstration of concept for the Met Office, we were asked to develop an emulator for HadCM3, based on 24 runs of the simulator, with a variety of parameter choices and future CO2 scenarios.

We had access to some runs of FAMOUS (a lower resolution model), which consisted of 6 scenarios for future CO2 forcing, and between 40 and 80 runs of FAMOUS under each scenario, with different parameter choices.

[And very little time to do the analysis.]

Design

Our design was

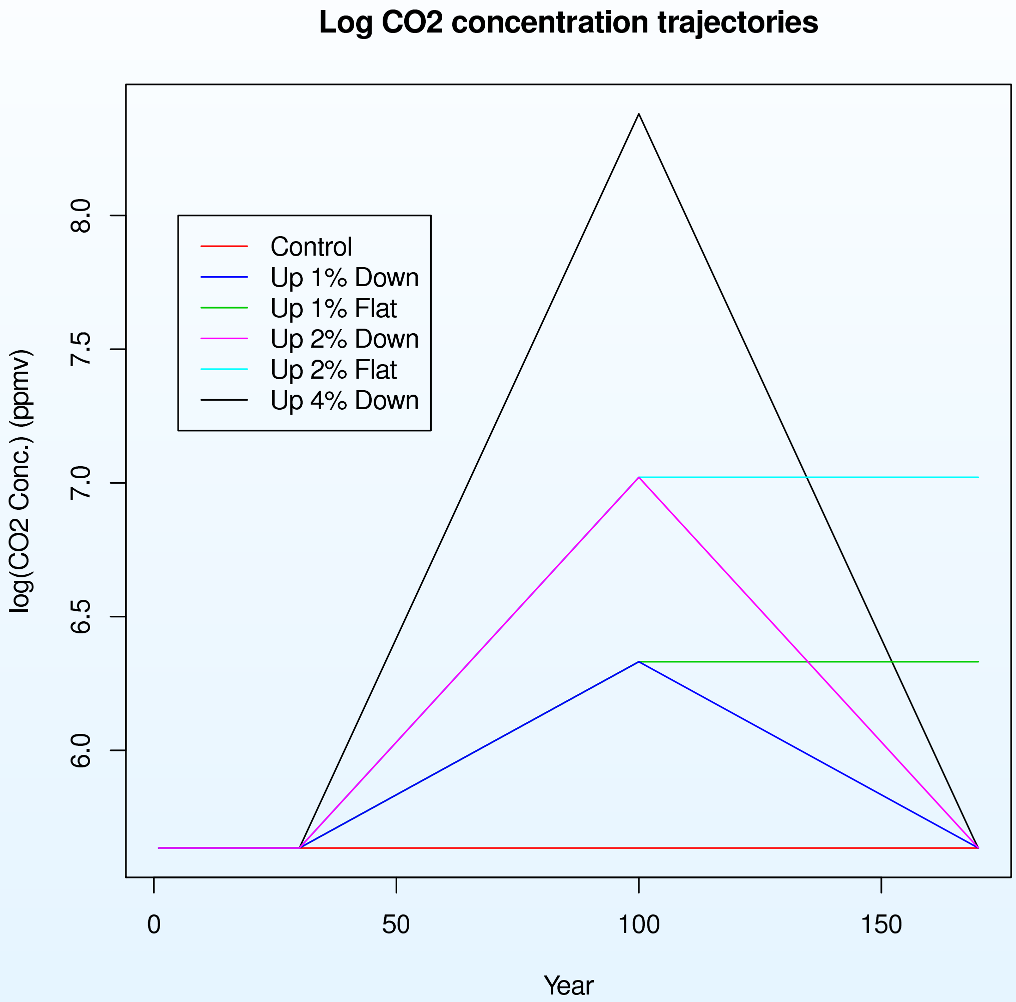
(i) to match the inputs for 8 of the HadCM3 runs with corresponding inputs to a FAMOUS run (to help us to compare the models)

(ii) to construct a 16 run Latin hypercube over different parameter choices and CO2 scenarios (to extend the model across CO2 space).

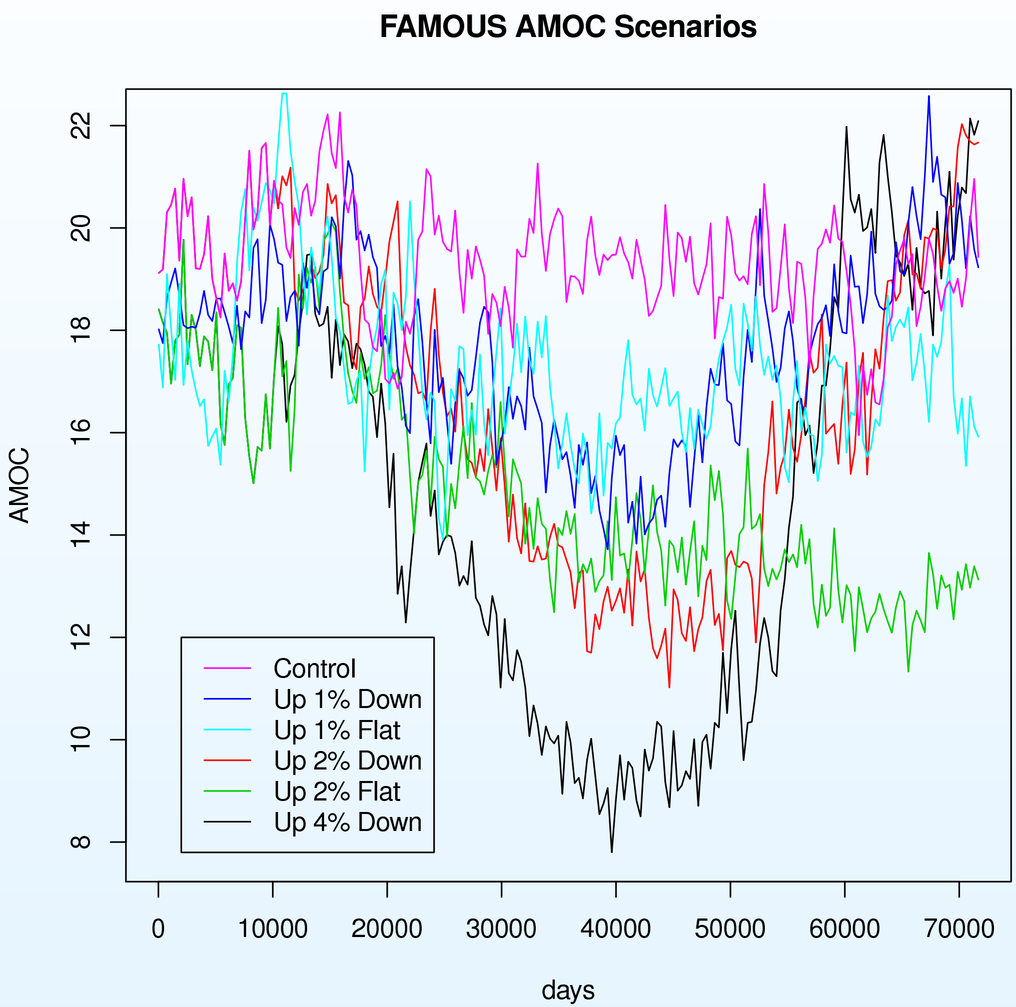
In this experiment only 3 parameters were varied (an entrainment coefficient in the model atmosphere, a vertical mixing parameter in the ocean, and the solar constant).

Our output of interest was a 170 year time series of AMOC values. The series is noisy and and the location and direction of spikes in the series was not important. Interest concerned aspects such as the value and location of the smoothed minimum of the series and the amount that AMOC responds to CO2 forcing and recovers if CO2 forcing is reduced.

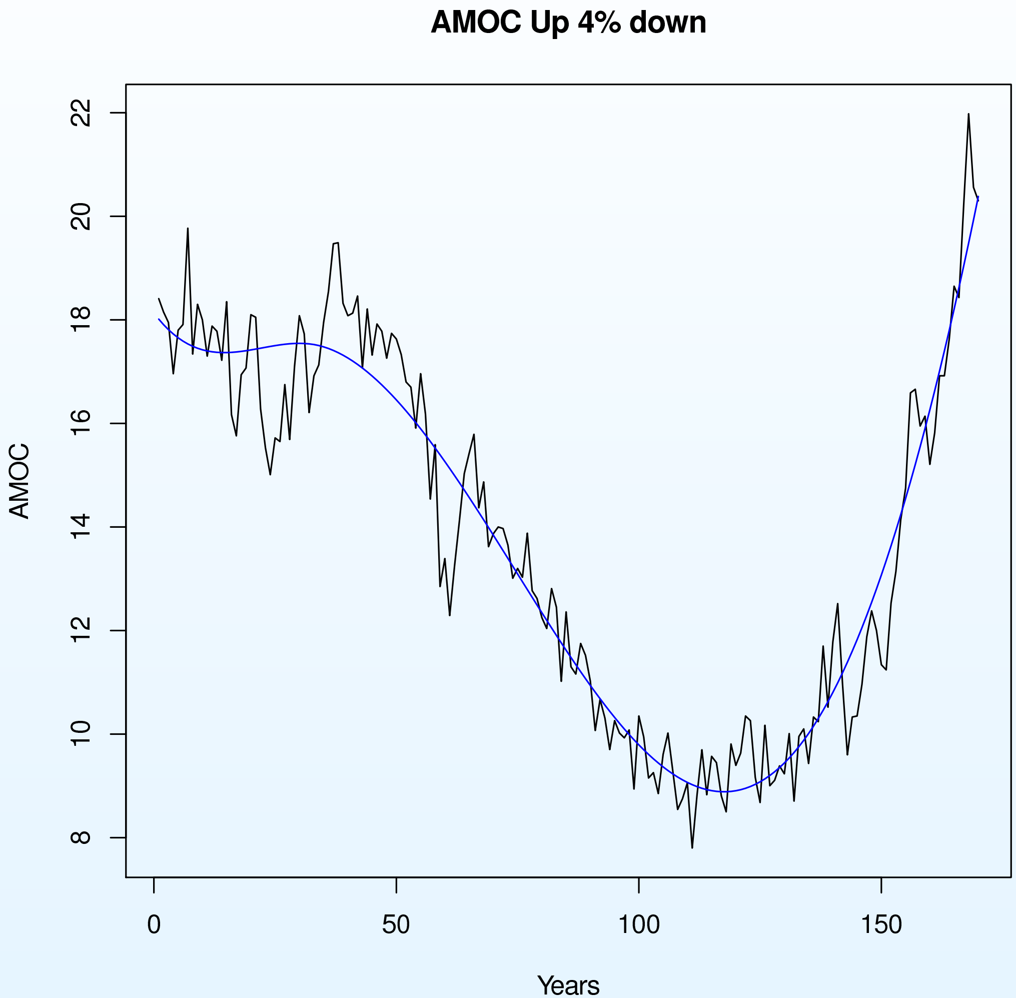
CO2 Scenarios



Famous Scenarios

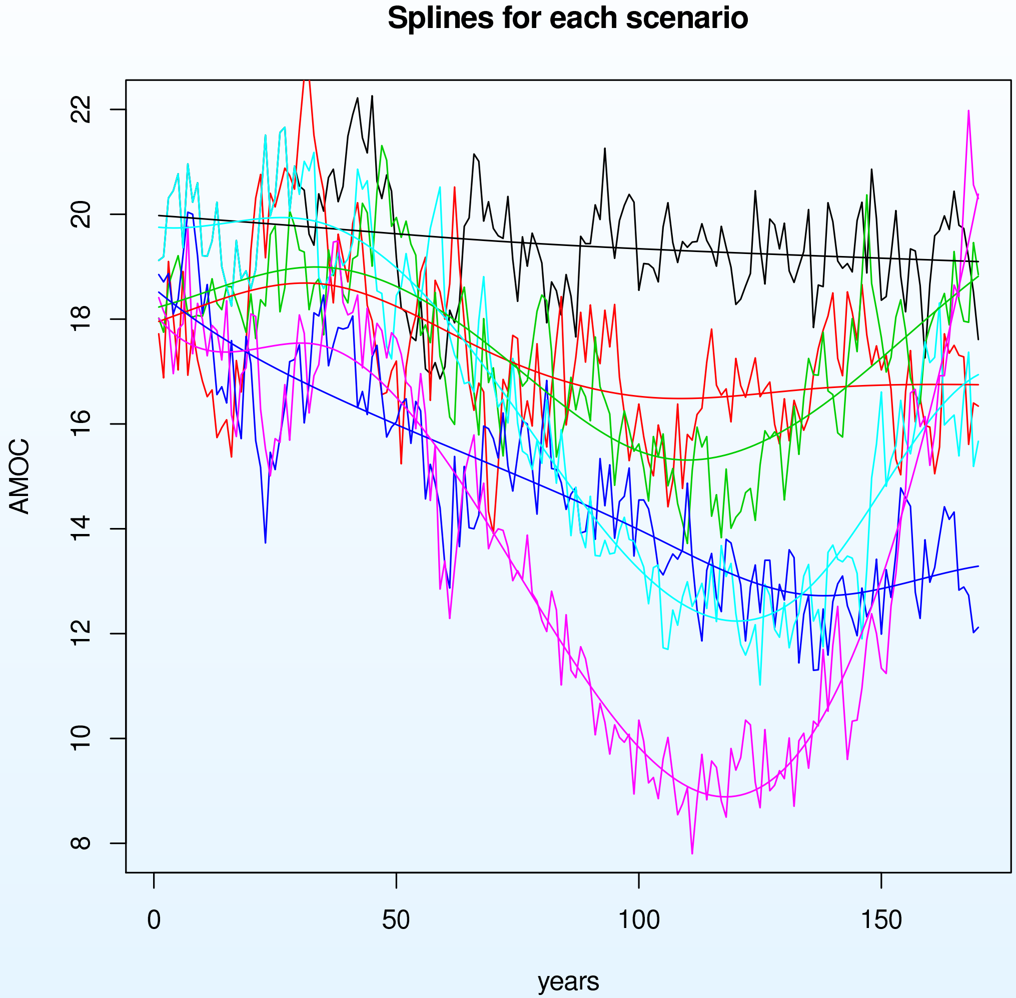


Smoothing



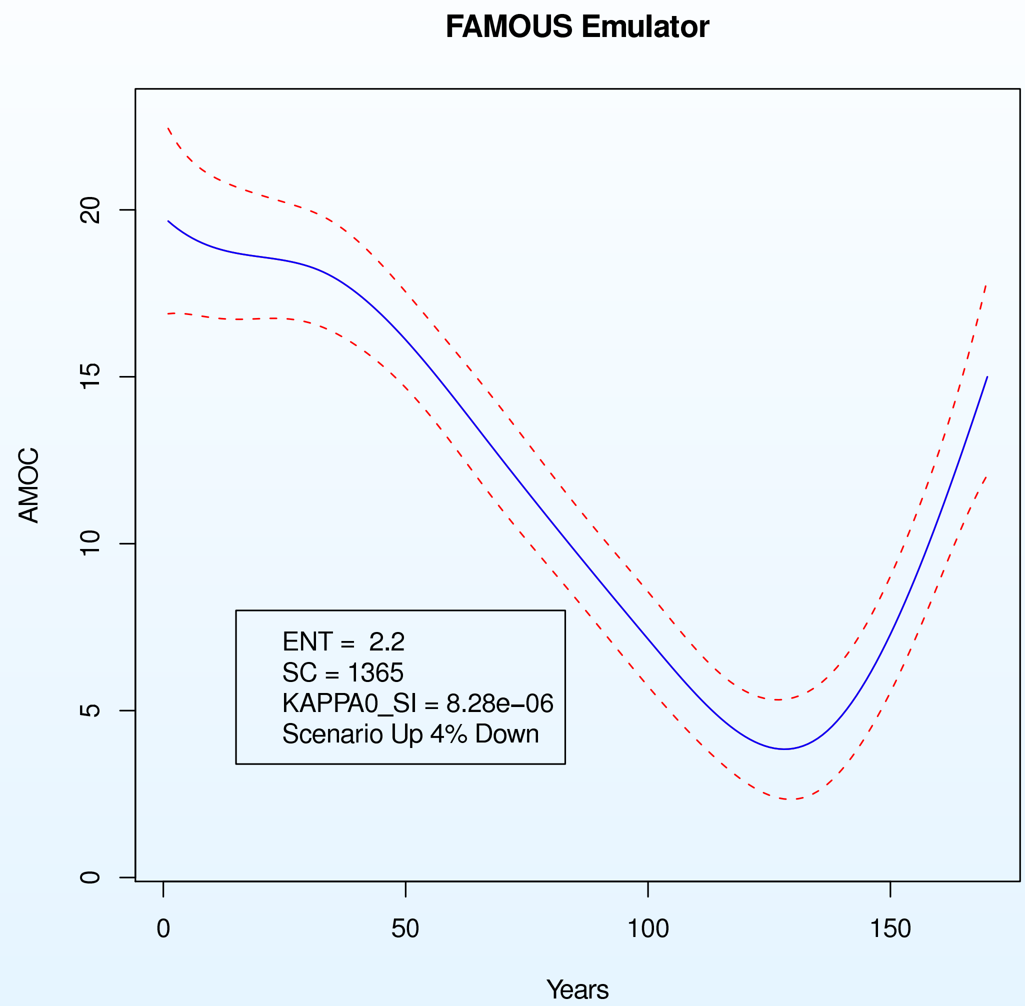
We smooth by fitting splines $f^s(x,t) = \Sigma_j c_j(x) B_j(t)$ where $B_j(t)$ are basis functions over t and $c_j(x)$ are chosen to give the 'best' smooth fit to the time series.

Smoothing



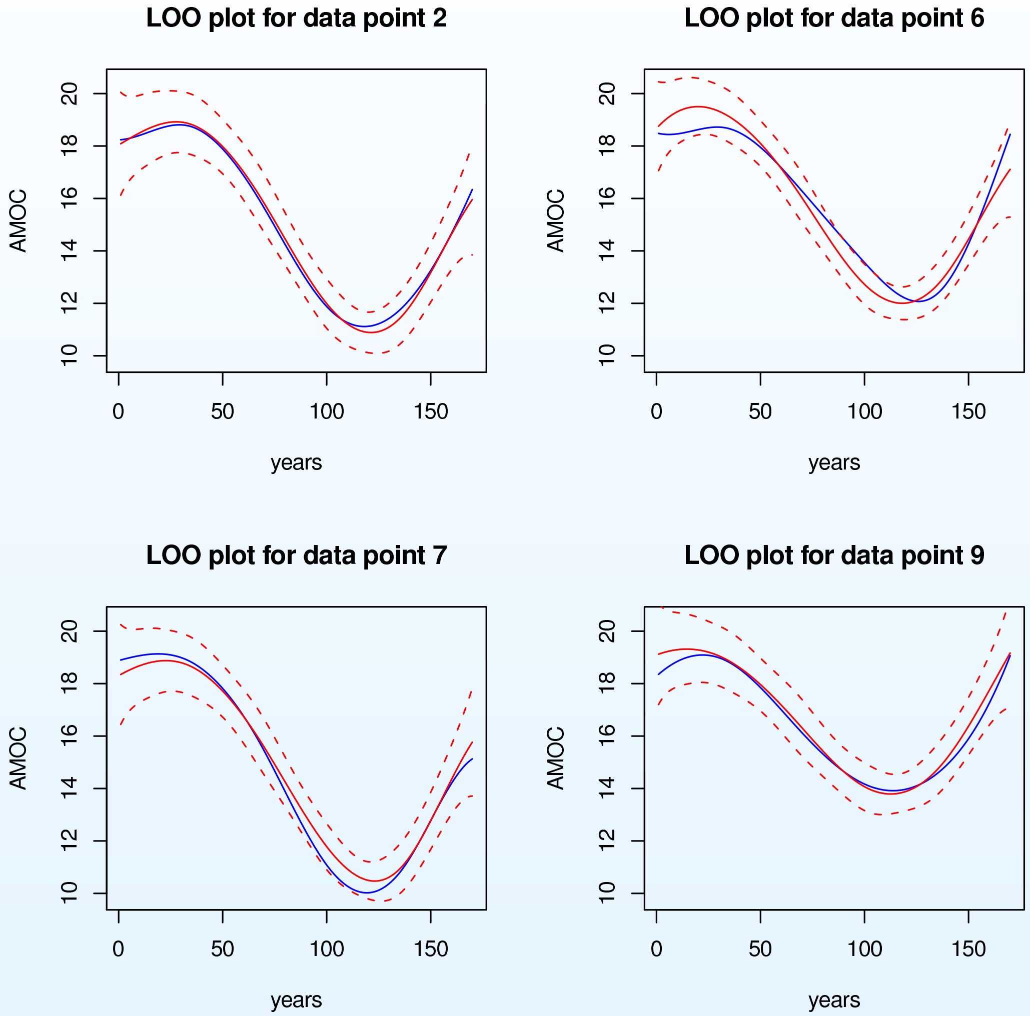
Splines for each scenario
AMOC
years
8
10
12
14
16
18
20
22
0
50
100
150

FAMOUS emulation



We emulate f^s by emulating each coefficient $c_j(x)$ in $f^s(x,t) = \Sigma_j c_j(x) B_j(t)$ (separately for each CO2 scenario)

Diagnostics (leave one out)



We test our approach by building emulators leaving out each observed run in turn, and checking whether the run falls within the stated uncertainty limits.

Emulating HadCM3

We now have an emulator for the smoothed version of FAMOUS, for each of the 6 CO_2 scenarios.

Next steps

[1] Extend the FAMOUS emulator across all choices of CO_2 scenario.

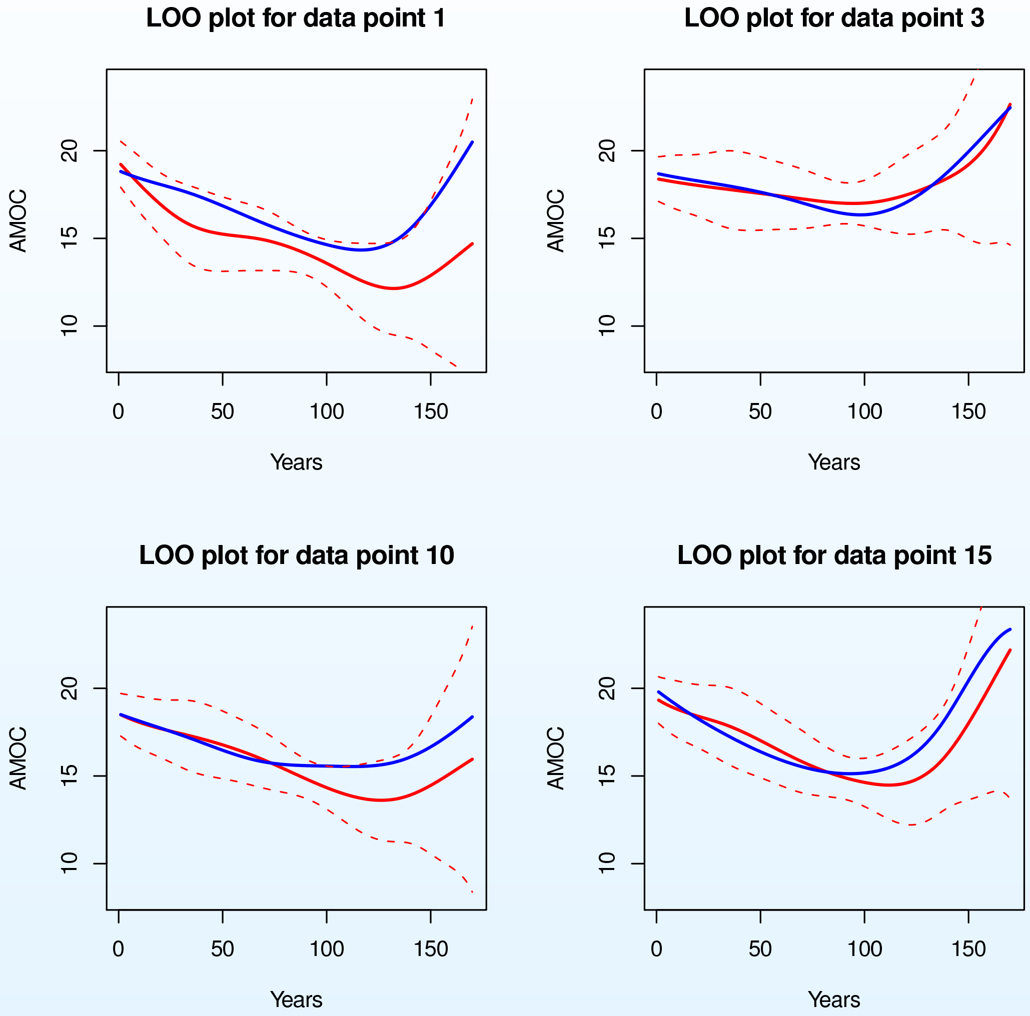
[We do this using fast geometric arguments, exploiting the speed of working in inner product spaces. For example, we have a different covariance matrix for local variation at each of 6 CO_2 scenarios. We extend this specification to all possible CO_2 scenarios by identifying each covariance matrix as an element of an appropriate inner product space, and adjusting beliefs over covariance matrix space by projection.]

[2] Develop relationships between the elements of the emulator for FAMOUS and the corresponding emulator for HadCM3, using the paired runs, and expert judgements. This gives an informed prior for the HadCM3 emulator.

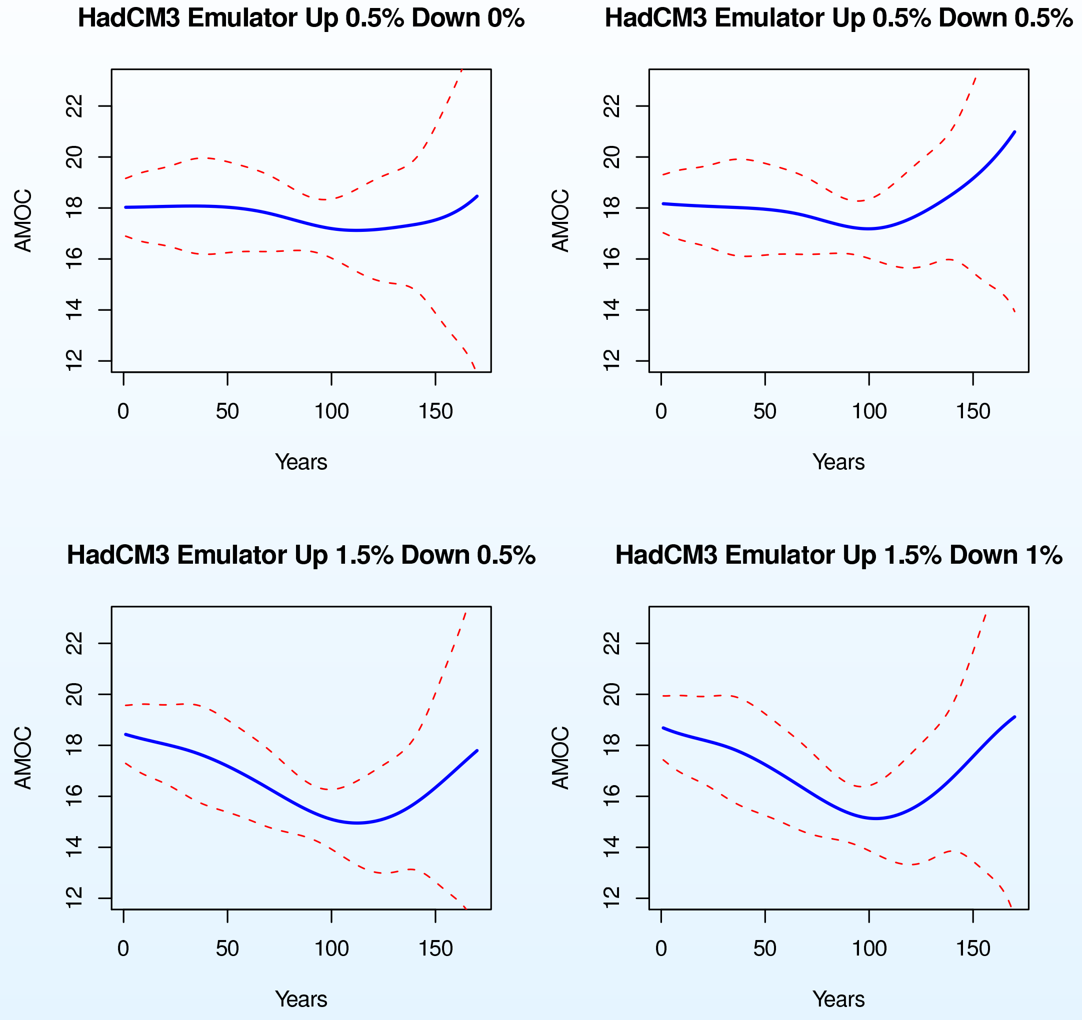
[3] Use the remaining runs of HadCM3 to Bayes linear update the emulator for HadCM3.

[4] Diagnostic checking, tuning etc.

Emulating HadCM3:diagnostics



Emulating HadCM3



Example: Oil Reservoirs

- A oil reservoir is an underground region of porous rock which contains oil and/or gas. The hydrocarbons are trapped above by a layer of impermeable rock and below by a body of water, thus creating the reservoir. The oil and gas are pumped out of the reservoir and fluids are pumped into the reservoir (to boost production).
- The simulator models the flows and distributions of contents of the reservoir over time
- Our Bayes linear approach to reservoir history matching has been implemented in software in use with Phillips-Conoco, Anadarko, Shell, Saudi-Aramco, ..
- Example - Oil field containing 650 wells, 1 million plus grid cells (permeability, porosity, fault lines, etc.). Previous history match took one man-year of effort. Our methods found a match using 32 runs, each lasting 4 hours and automatically chosen with a overall fourfold improvement in fit.

Inputs and outputs

Inputs

Each cell in the reservoir has a collection of associated input parameters, such as permeability and porosity. There are also other parameters, such as Fault transmissibility, Aquifer features, Saturation properties

Since there are a huge number of these cells in the reservoir it is common to use scalar multipliers over subregions, to modify values.

Outputs

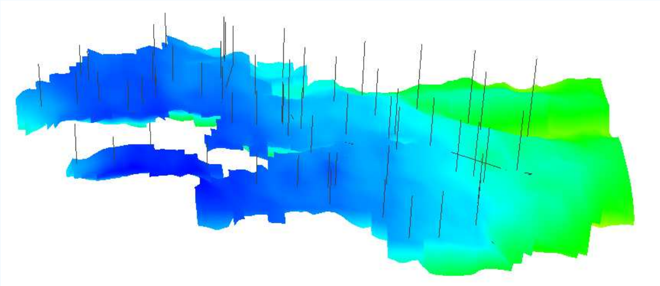
- The model outputs comprise the behaviour of the various wells and injectors in the reservoir
- Output is a time series on the following variables for each well
  - *Pressures* Bottom-hole pressure, Tubing head pressure
  - *Production/Injection rates and totals* for each of oil, water and gas.
  - *Fluid ratios* Water cut, Gas-oil ratio
- The resolution of the time series can be varied from months to years
- With a large number of wells, daily output, or a long operating period there will be a *lot* of output data

A reservoir example: (thanks to Jonathan Cumming)

The model, based on grid size $38 \times 87 \times 25$, with 43 production and 13 injection wells, simulates 10 years of production, 1.5–3 hours per simulation.

Inputs Field multipliers for porosity (ϕ), permeabilities (k_x, k_z), critical saturation (crw), and aquifer properties (A_p, A_h)

Outputs Oil production rate for a 3-year period, for the 10 production wells active in that period. 4-month averages over the time series



Emulator for reservoir simulator is: $f_i(\boldsymbol{x}) = \boldsymbol{g}_i(\boldsymbol{x}_{[i]})\boldsymbol{\beta}_i + u_i(\boldsymbol{x}_{[i]}) + v_i(\boldsymbol{x})$

$\boldsymbol{g}_i(\boldsymbol{x}_{[i]})^T\boldsymbol{\beta}_i$ – a global trend function which captures the gross features, $\boldsymbol{x}_{[i]}$ – a subset of inputs which account for most of the variation in F, the *active variables*, $u_i(\boldsymbol{x}_{[i]})$ – a correlated residual process representing the local behaviour in the active variables, $v_i(\boldsymbol{x})$ – an uncorrelated 'nugget' residual.

Coarse and Accurate Emulators

The computer model is expensive to evaluate, so we use ‘coarse’ model, F^c,to capture qualitative features of F. F^c is substantially faster, allowing many model runs. We construct emulator f^c of F^c from these runs and a framework linking f^c and f. We make (small) number of runs of F, and update our emulator f.

Obtain F^c by coarsening vertical gridding by factor of 10.

1000 runs of F^c in a Latin Hypercube over the input parameters

Screen the wells (Principal Variables methods) – 4 wells capture 87% of the total variation

We fit emulators to each output individually, using information from the model runs (stepwise regression and generalised least squares) to get emulator $f_i^c(\boldsymbol{x})$ for F_i^c.

We consider the coarse and the full model emulators to have the form
$f_i^c(\boldsymbol{x}) = \boldsymbol{g}_i(\boldsymbol{x}_{[i]})^T \boldsymbol{\beta}_i^c + w_i^c(\boldsymbol{x}), f_i(\boldsymbol{x}) = \boldsymbol{g}_i(\boldsymbol{x}_{[i]})^T \beta_i + w_i^c(\boldsymbol{x})\beta_{w_i} + w_i^a(\boldsymbol{x})$
(linked via the equations relating the pairs of coefficients)

Careful choice of small design to evaluate for full simulator allows us to (Bayes linear) update emulator for F based on prior emulator and additional runs.

Emulation Summaries

Well	Time	$\boldsymbol{x}_{[i]}$	No. Model Terms	Coarse Simulator R^2	Accurate Simulator $\tilde{R}^2$
B2	4	ϕ, crw, A_p	9	0.886	0.951
B2	8	ϕ, crw, A_p	7	0.959	0.958
B2	12	ϕ, crw, A_p	10	0.978	0.995
B2	16	ϕ, crw, k_z	7	0.970	0.995
B2	20	ϕ, crw, k_x	11	0.967	0.986
B2	24	ϕ, crw, k_x	10	0.970	0.970
B2	28	ϕ, crw, k_x	10	0.975	0.981
B2	32	ϕ, crw, k_x	11	0.980	0.951
B2	36	ϕ, crw, k_x	11	0.983	0.967

History matching

Model calibration aims to identify “true” input parameters x^*. However

(i) We may not believe in a unique true input value for the model;

(ii) We may be unsure whether there are any good choices of input parameters (due to model deficiencies)

(iii) Full Bayes calibration analysis may be very difficult/non-robust.

A conceptually simple alternative is “history matching”, i.e. finding the collection of all input choices x for which you judge the match of the model to the data, $\|z - f_h(x)\|$ to be acceptably small, using some ”‘implausibility measure”' $I(x)$ based on a natural probabilistic metric, accounting for emulator uncertainty, condition uncertain, structural discrepancy, observational error etc.

In practice, we proceed by sequentially ruling out regions of x space which are unlikely to give rise to observed history z.

History matching via Implausibility

Using the emulator we can obtain, for each set of inputs x, the mean and variance, $\mathrm{E}(F_h(x))$ and $\mathrm{Var}(F_h(x))$.

As $z_i = y_i + e_i,\ y_i = F_i^* + \epsilon_i$,
if $x = x^*$, then
$\mathrm{Var}(z_i - \mathrm{E}(F_i(x))) = \mathrm{Var}(F_i(x)) + \mathrm{Var}(\epsilon_i) + \mathrm{Var}(e_i)$.

We can therefore calculate, for each output $F_i(x)$, the "implausibility" if we consider the value x to be the best choice x^*, which is the standardised distance between z_i and $\mathrm{E}(F_i(x))$, which is

$$I_{(i)}(x) = |z_i - \mathrm{E}(F_i(x))|^2 / [\mathrm{Var}(F_i(x)) + \mathrm{Var}(\epsilon_i) + \mathrm{Var}(e_i)]$$

[Large values of $I_{(i)}(x)$ suggest that it is implausible that $x = x^*$.]

Using Implausibility measures

The implausibility calculation can be performed univariately, or by multivariate calculation over sub-vectors. The implausibilities are then combined, such as by using $I_M(x) = \max_i I_{(i)}(x)$, and can then be used to identify regions of x with large $I_M(x)$ as implausible, i.e. unlikely to be good choices for x^*.

With this information, we can then refocus our analysis on the ‘non-implausible’ regions of the input space, by

(i) making more simulator runs

(ii) refitting our emulator

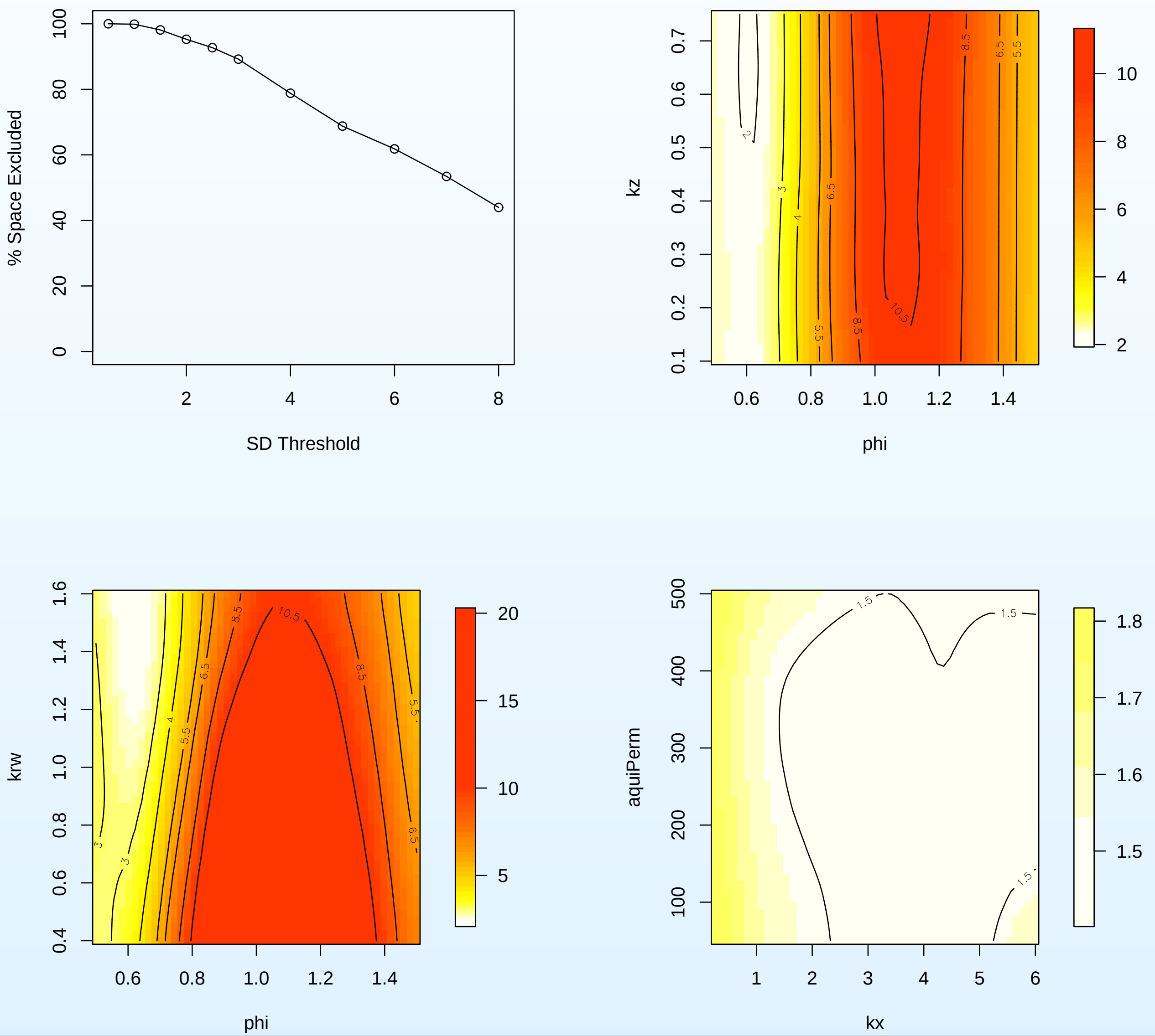
over such sub-regions and repeating the analysis.

This process is a form of iterative global search aimed at finding all choices of x^* which would give good fits to historical data.

Comment We may find no good choices at all which give good fits and that is a clear sign of problems with our physical simulator or with our data.

Comment: Even if calibrating, it is good practice to history match first, to check model and (massively) reduce search space.

Implausibility Results



Refocusing

- Make the restriction $\mathcal{X}^* = \{\boldsymbol{x} : \mathcal{I}(\boldsymbol{x}) \leq 4\} \simeq \{\boldsymbol{x} : \phi < 0.79\}$ and eliminate 90% of the input space
- Now consider final 4 time points in original data, plus an additional point 1 year beyond the end of the previous series to be forecast
- Since reducing the space many of the old model runs are no longer valid, so supplement with additional evaluations
- 262+100 coarse runs, 6+20 accurate runs
- Re-fit the coarse and fine emulators, using the old emulator structure as a starting point

Forecasting

The mean and variance of $F(x)$ are obtained from the mean function and variance function of the emulator f for F. Using these values, we compute the mean and variance of $F^* = F(x^*)$ by first conditioning on x^* and then integrating out x^*.

Given $\mathrm{E}(F^*)$, $\mathrm{Var}(F^*)$, and the model discrepancy, ϵ and sampling error e variances, it is now straightforward to compute the joint mean and variance of the collection (y, z) (as $y = F^* + \epsilon, z = y + e$).

We now evaluate the adjusted mean and variance for y_p adjusted by z using the Bayes linear adjustment formulae. This analysis is tractable even for large systems.

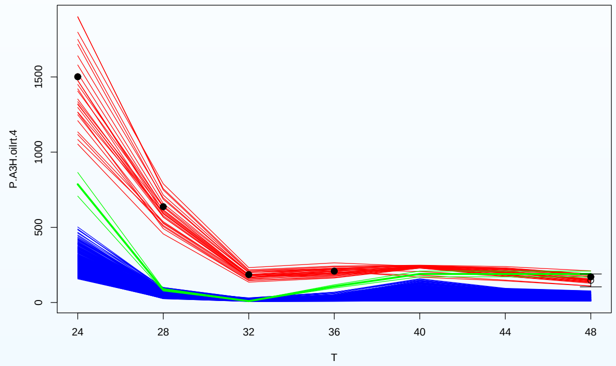
(When the forecast variance is large, then we have methods to improve forecast accuracy.)

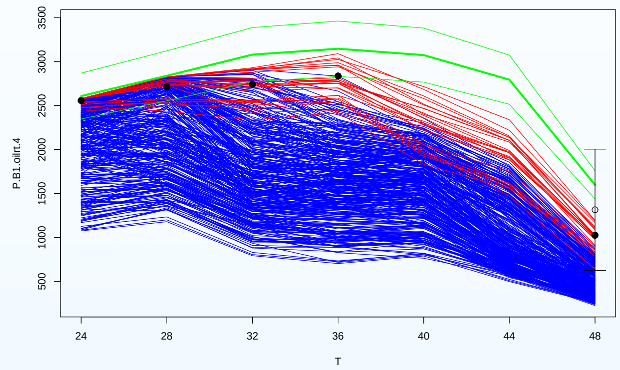
Comment Our computer experiments to forecast y_p split into two stages

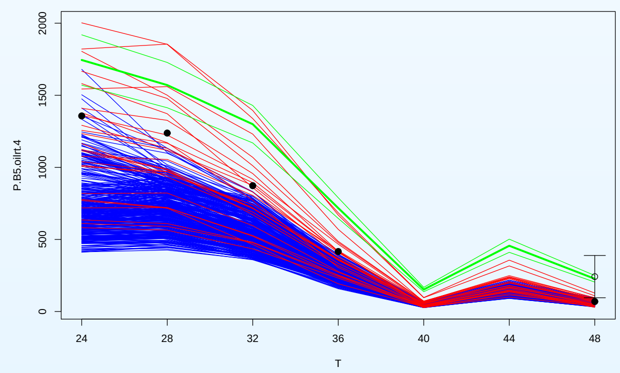
(i) preliminary simulator evaluations to identify the form of emulator, estimate coefficient matrices and refocus

(ii) further simulator evaluations chosen to minimise adjusted forecast variance.

Forecasting Results







Simulator outputs, observational data and forecasts for each well.

Green lines indicate z with error bounds of 2sd(e).

Red and blue lines represent the range of the runs of $F(x)$ and $F^c(x)$

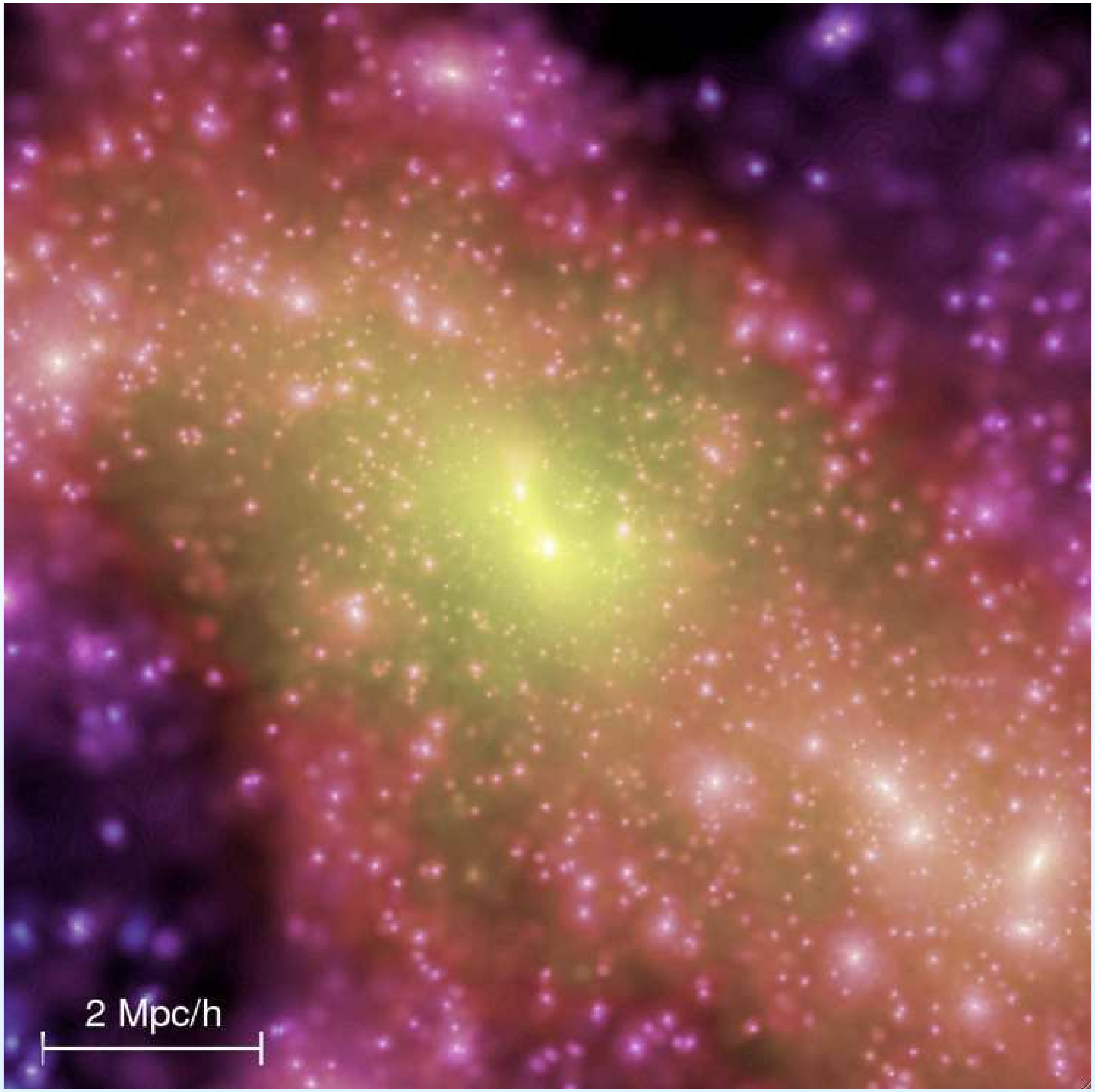
Solid black dots correspond to $\mathrm{E}(F^*)$.

The forecast is indicated by a hollow circle with attached error bars.

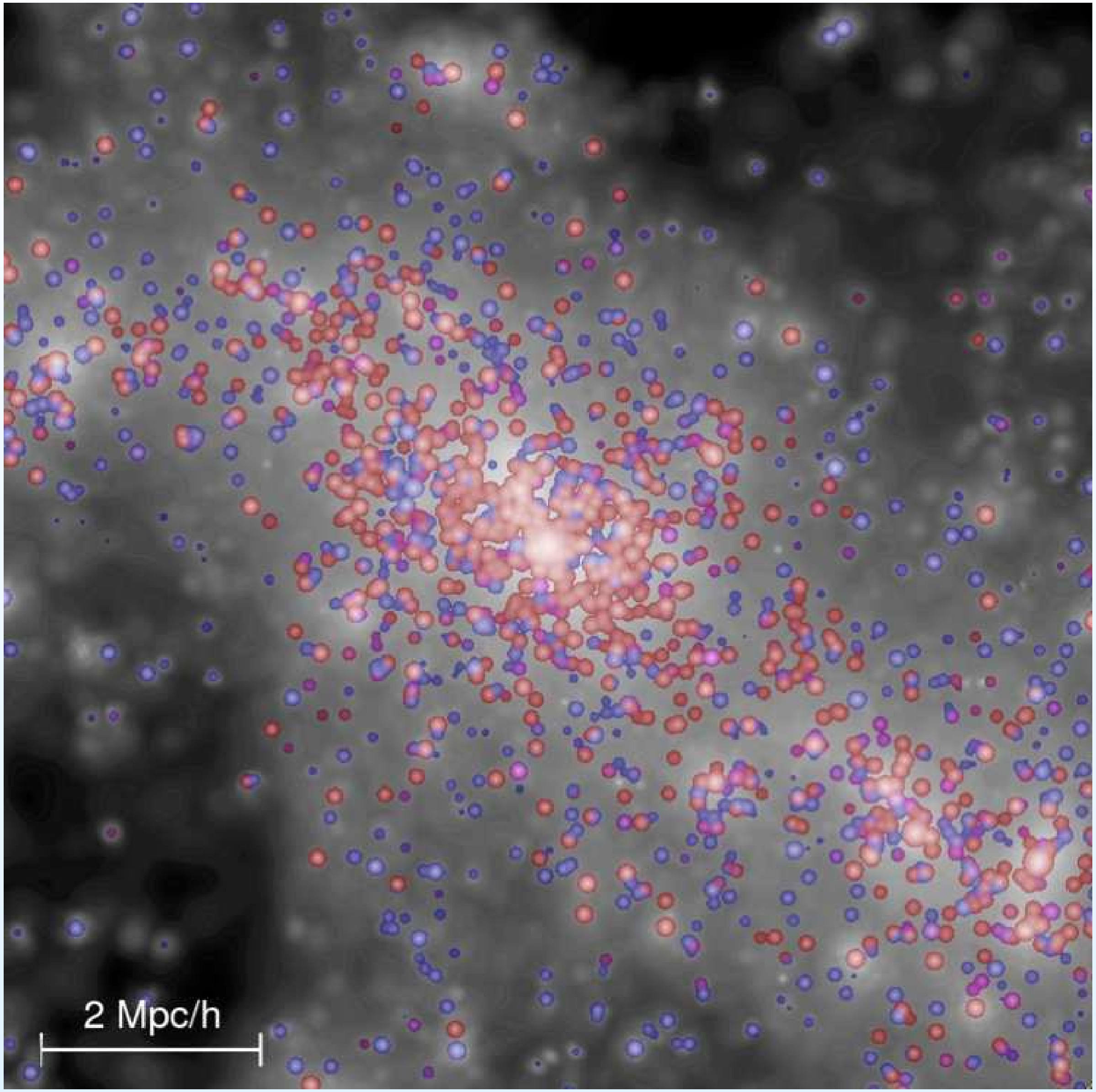
The Galform Model (thanks to Ian Vernon)

- The Cosmologists at the ICC are interested in modelling galaxy formation in the presence of Dark Matter.
- First a Dark Matter simulation is performed over a volume of (1.63 billion light years$)^3$. This takes 3 months on a supercomputer.
- Galform takes the results of this simulation and models the evolution and attributes of approximately 1 million galaxies.
- Galform requires the specification of 17 unknown inputs in order to run.
- It takes approximately 1 day to complete 1 run (using a single processor).
- The Galform model produces lots of outputs, some of which can be compared to observed data from the real Universe.

The Dark Matter Simulation



The Galform Model



Inputs

To perform one run, we need to specify the following 17 inputs:

vhotdisk:	100 - 550	**VCUT**:	20 - 50
aReheat:	0.2 - 1.2	**ZCUT**:	6 - 9
alphacool:	0.2 - 1.2	**alphastar**:	-3.2 - -0.3
vhotburst:	100 - 550	**tau0mrg**:	0.8 - 2.7
epsilonStar:	0.001 - 0.1	**fellip**:	0.1 - 0.35
stabledisk:	0.65 - 0.95	**fburst**:	0.01 - 0.15
alphahot:	2 - 3.7	**FSMBH**:	0.001 - 0.01
yield:	0.02 - 0.05	**eSMBH**:	0.004 - 0.05
tdisk:	0 - 1		

Galform provides multiple output data sets. Initially we analyse luminosity functions giving the number of galaxies per unit volume, for each luminosity.

Bj Luminosity: corresponds to density of young (blue) galaxies

K Luminosity: corresponds to density of old (red) galaxies

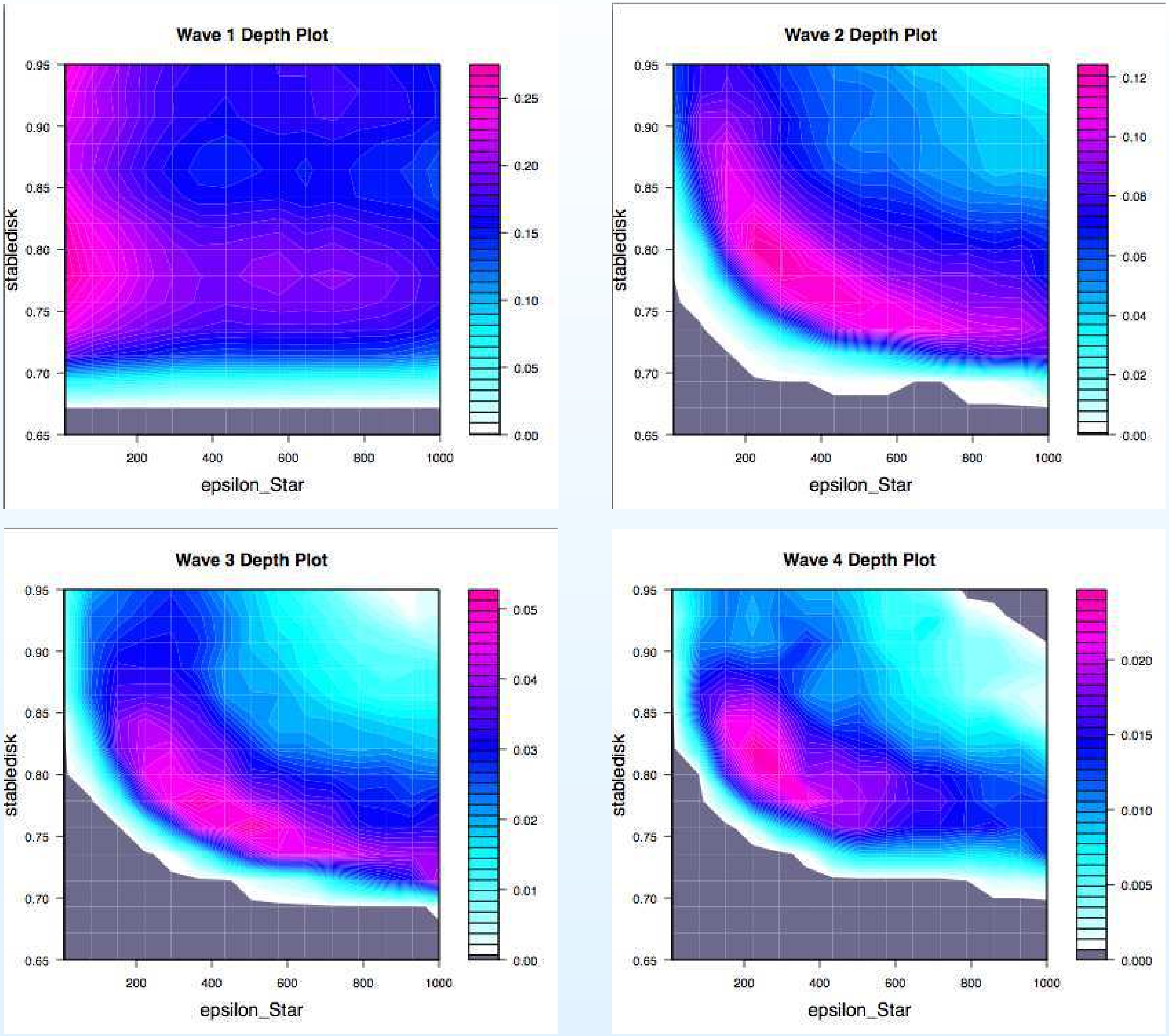
We choose 11 outputs that are representative of the Luminosity functions and emulate the functions $f_i(x)$.

Summary of Results

We assess condition uncertainty, structural uncertainty, measurement uncertainty, etc. then carry out the iterative history matching procedure, through 4 waves.
(In wave 5, we evaluate many good fits to data, and we stop. Some of these choices give simultaneous matches to data sets that the Cosmologists have been unable to match before.)

	No. Model Runs	No. Active Vars	Space Remaining
Wave 1	1000	5	14.9 %
Wave 2	1414	8	5.9 %
Wave 3	1620	8	1.6 %
Wave 4	2011	10	0.12 %

2D Implausibility Projections: Wave 1 (14%) to Wave 4 (0.12%)



Linking models to reality

The reason that the evaluations of the simulator are informative for the physical system is that the evaluations are informative about the general relationships between system properties, x, and system behaviour y.

More generally, evaluations of a collection of models are jointly informative for the physical system as they are jointly informative for these general relationships.

Therefore, our inference from model to reality should proceed in two parts.

[1] We emulate the relationship between system properties and system behaviour (we call this relationship the “reified model” (from reify: to treat an abstract concept as if it were real).

[2] We decompose the difference between our model and the physical system into two parts.

[A] The difference between our simulator and the reified form.

[B] The difference between the reified form at the physically appropriate choice of x and the actual system behaviour y.

Relating models and the system

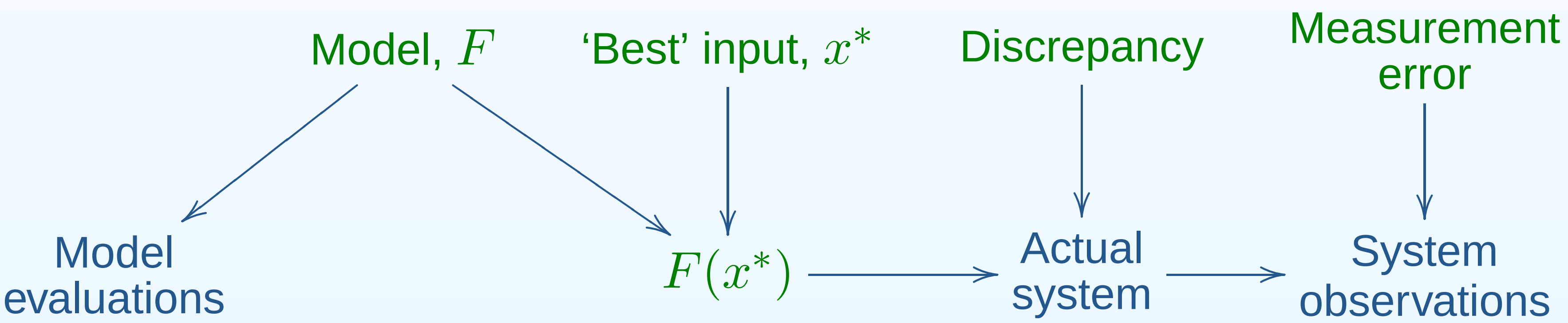
Reifying principle [1]

Simulator F is informative for y, because F is informative for F^* and $F^*(x^*)$ is informative for y.

Relating models and the system

Reifying principle [1]

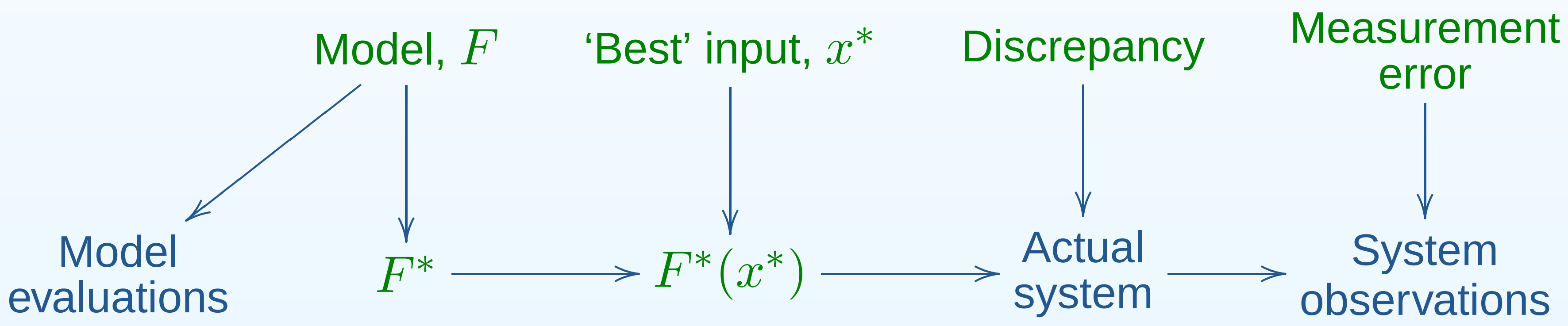
Simulator F is informative for y, because F is informative for F^* and $F^*(x^*)$ is informative for y.



Relating models and the system

Reifying principle [1]

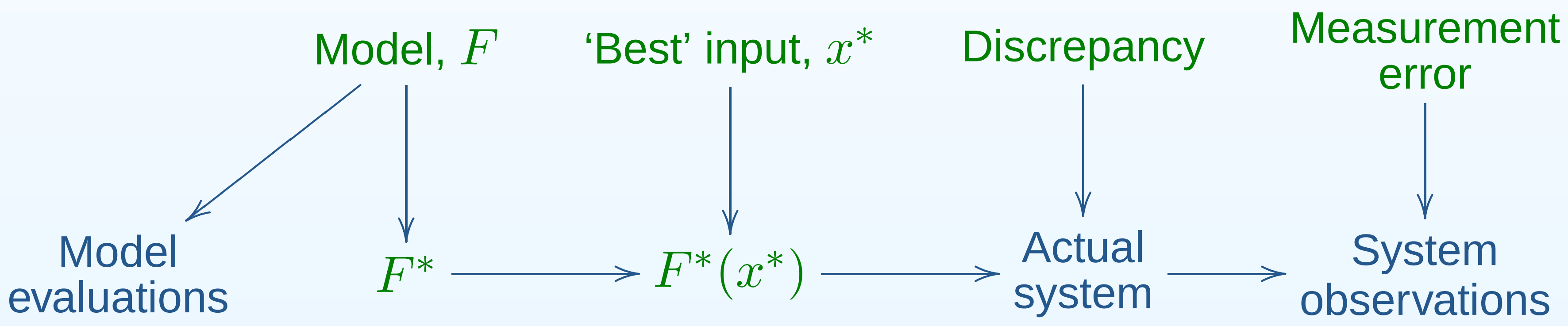
Simulator F is informative for y, because F is informative for F^* and $F^*(x^*)$ is informative for y.



Relating models and the system

Reifying principle [1]

Simulator F is informative for y, because F is informative for F^* and $F^*(x^*)$ is informative for y.



Reifying principle [2]

A collection of simulators $F_1, F_2, \ldots$ is jointly informative for y, as the simulators are jointly informative for F^*.

Linking F and F^* using emulators

Suppose that our emulator for F is

$$f(x) = Bg(x) \oplus u(x)$$

Our simplest emulator for F^* might be

$$f^*(x, w) = B^* g(x) \oplus u^*(x) \oplus u^*(x, w)$$

where we might model our judgements as $B^* = CB + \Gamma$, correlate $u(x)$ and $u^*(x)$, while $u^*(x, w)$, with additional parameters, w, is uncorrelated with remainder.

Structured reification improves on this with systematic modelling for all aspects of model deficiency whose effects we can consider explicitly.

All our calibration and forecasting methodology is unchanged - all that has changed is our description of the joint covariance structure.

A Reified influence diagram

$$\left[F^1_{h:[n]}(x), \ldots, F^m_{h:[n]}(x)\right]$$

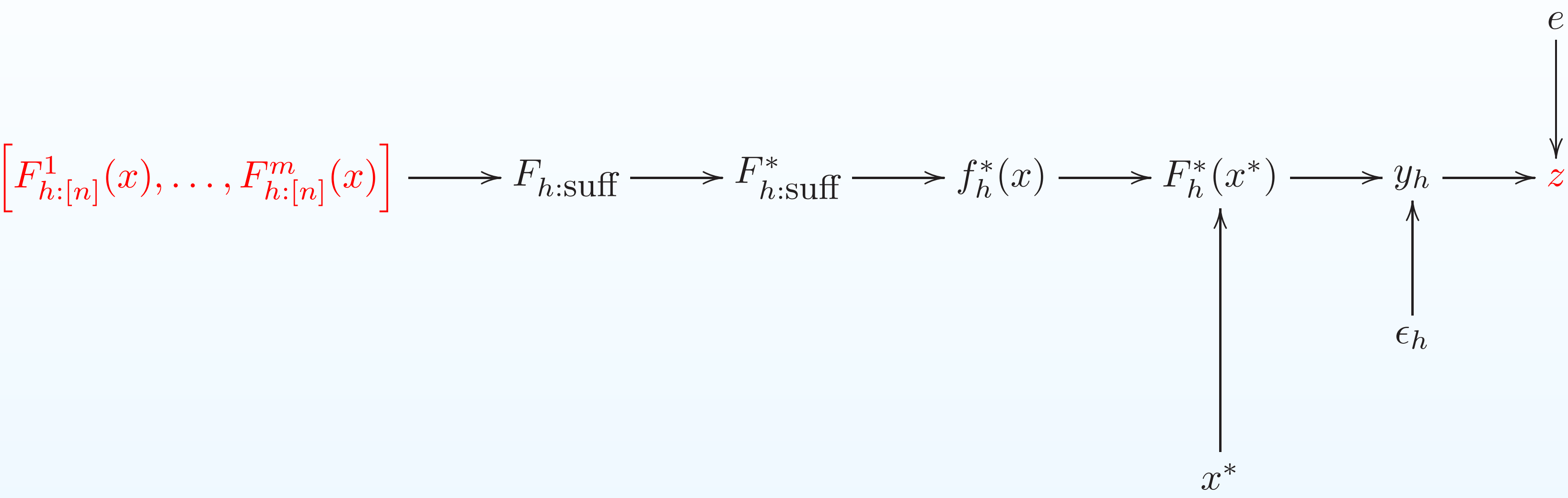
Evaluations of the simulator at each of m initial conditions for historical components of simulator

A Reified influence diagram

$$\left[F^1_{h:[n]}(x), \ldots, F^m_{h:[n]}(x)\right] \longrightarrow F_{h:\text{suff}} \longrightarrow F^*_{h:\text{suff}} \longrightarrow f^*_h(x)$$

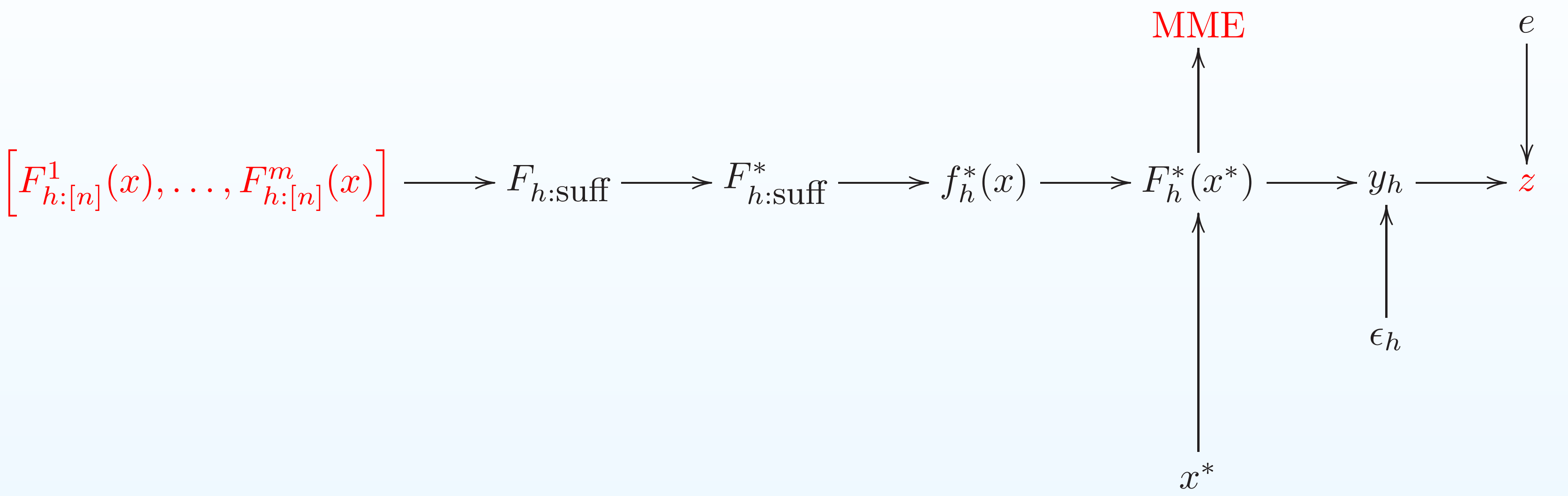
Global information $F_{\text{h:suff}}$ (from second order exchangeability modelling).
passes to Reified global form and to reified emulator.

A Reified influence diagram



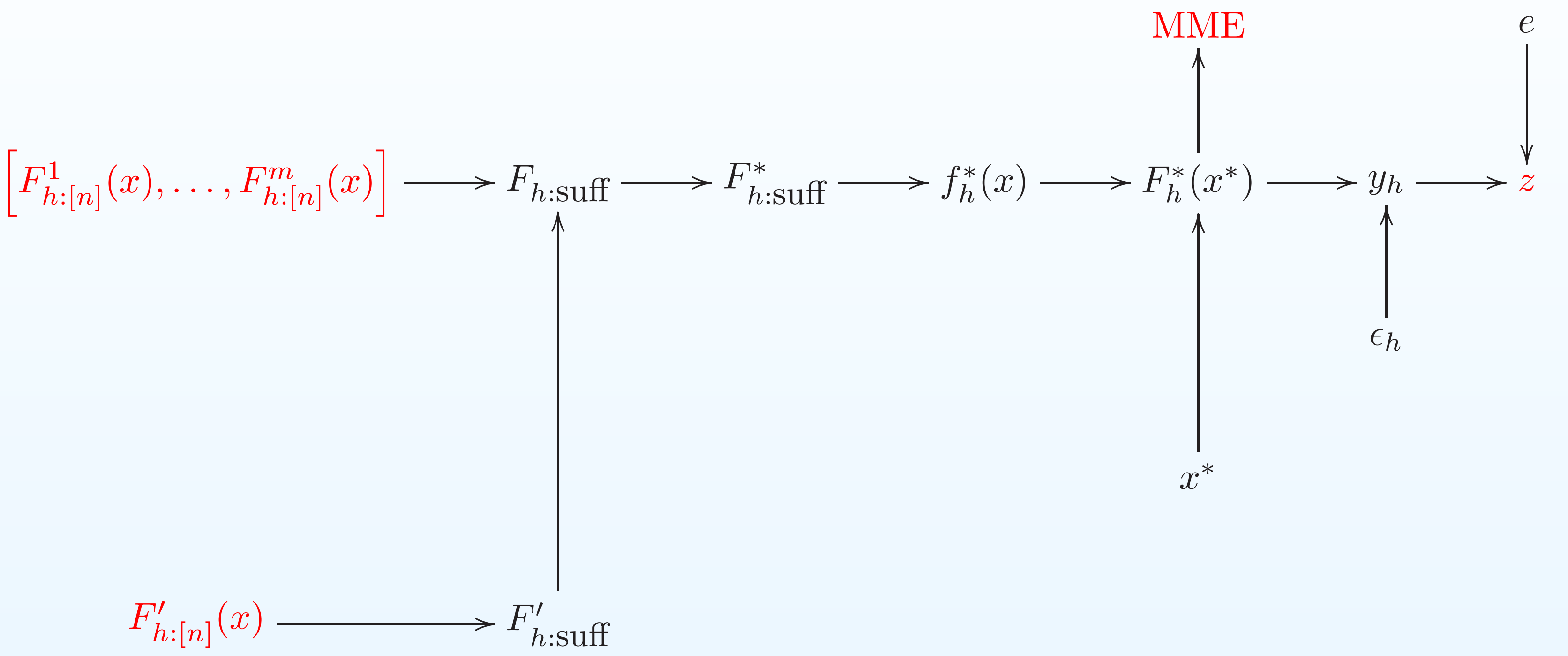
Link with x^* to reified function, at true initial condition, linked to data z

A Reified influence diagram



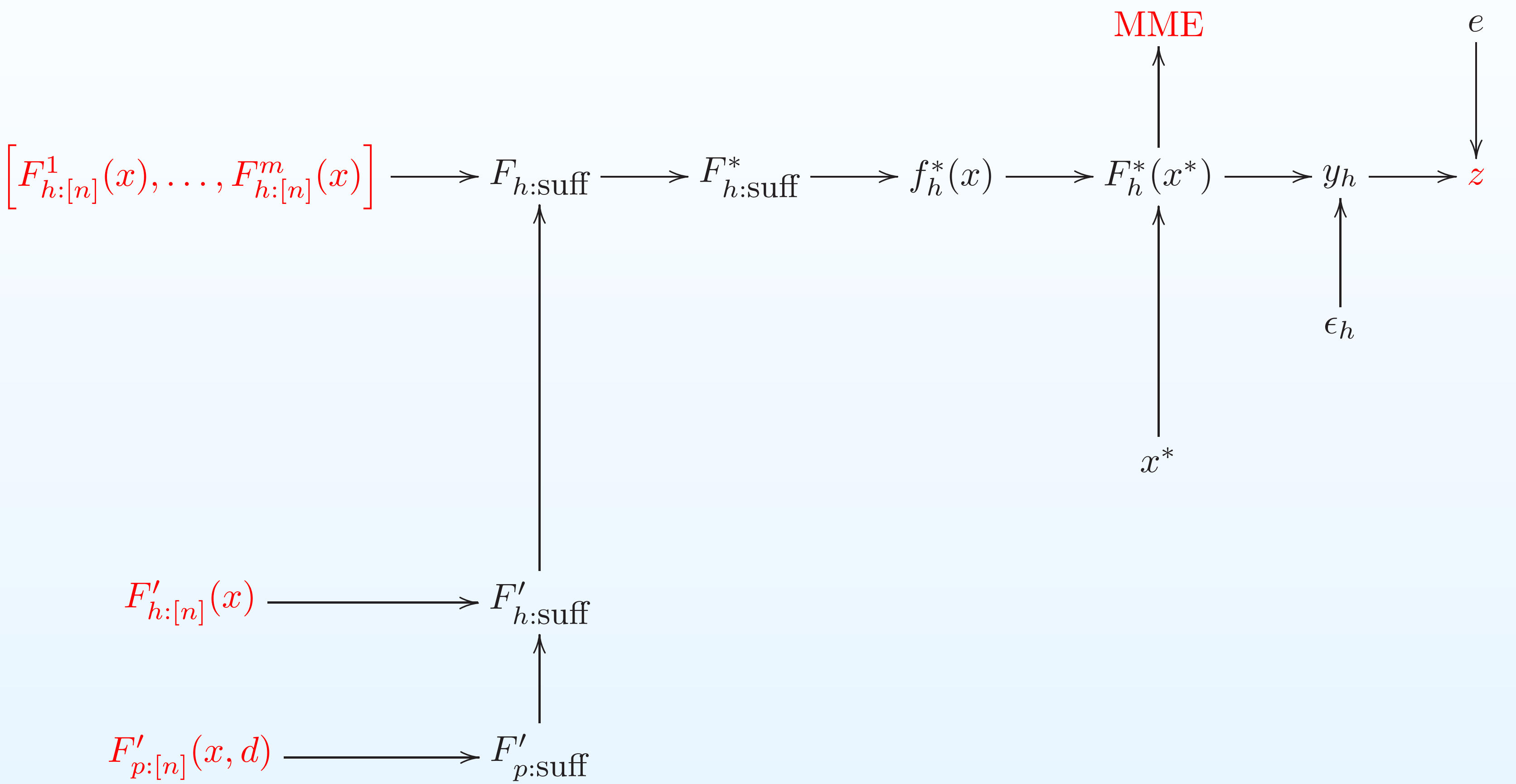
Add observation of a related multi-model ensemble (MME) consisting of tuned runs from related models (more exchangeability modelling).

A Reified influence diagram



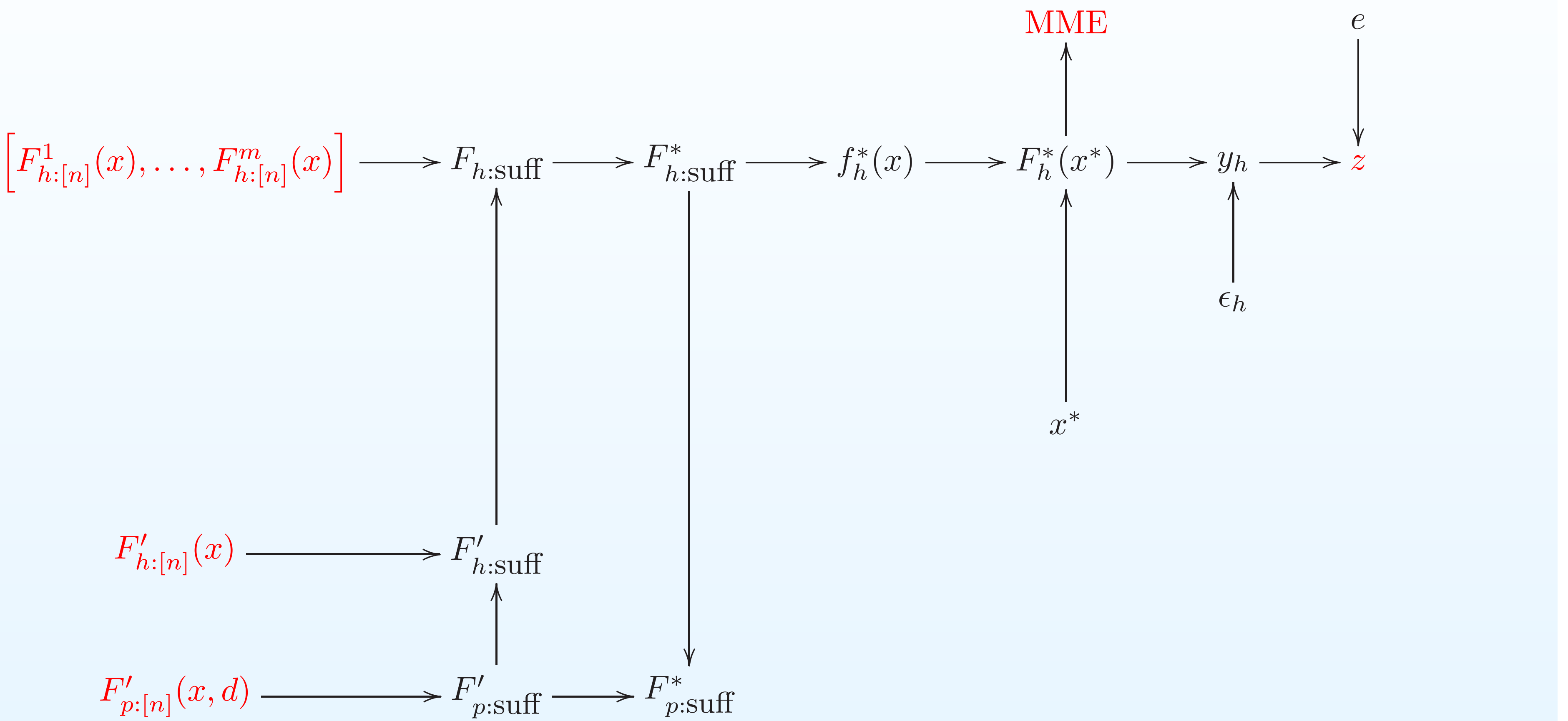
Add a set of evaluations from a fast approximation

A Reified influence diagram



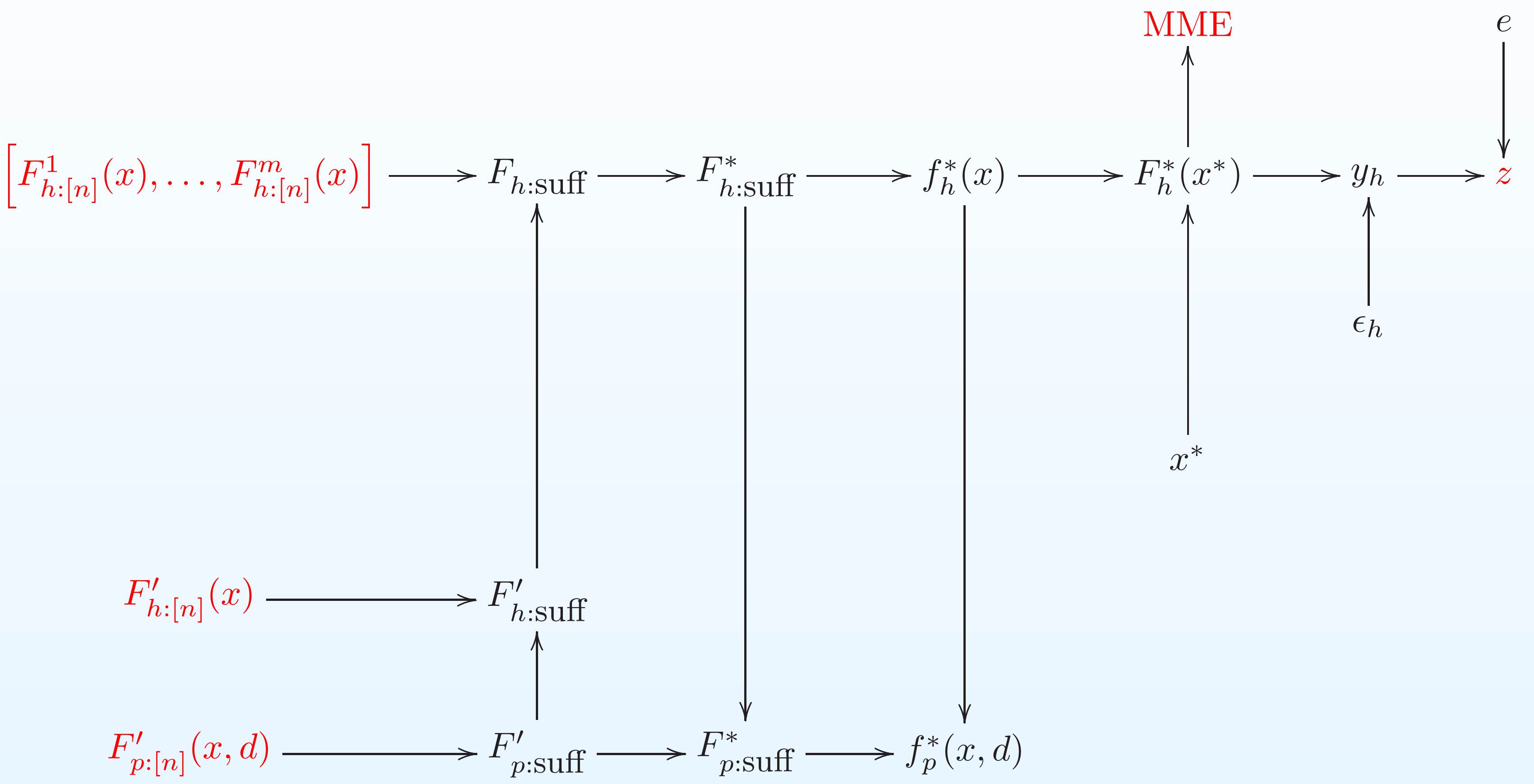
Add evaluations of fast simulator for outcomes to be predicted, with decision choices d

A Reified influence diagram



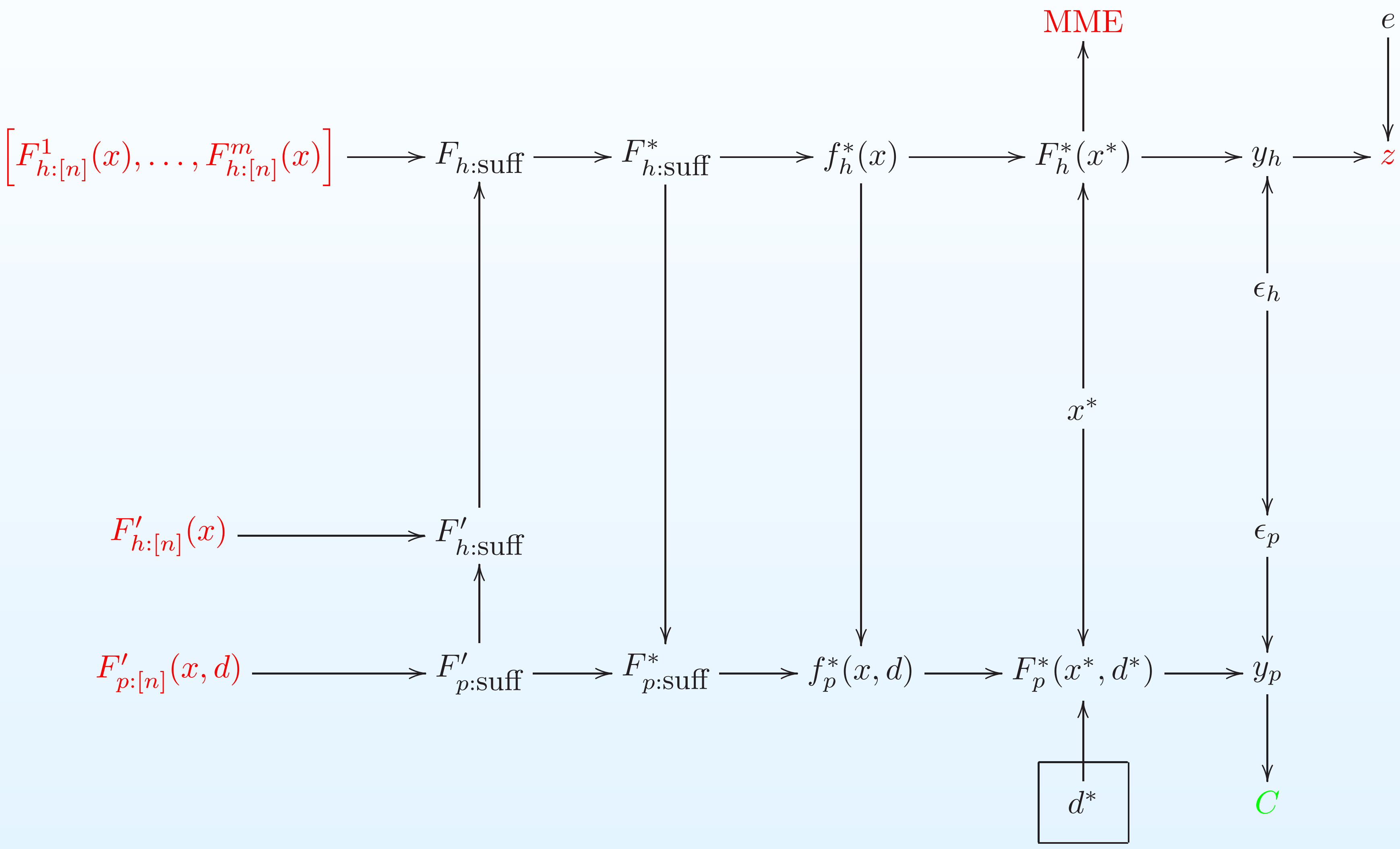
Link to reified global terms for quantities to be predicted

A Reified influence diagram



And to reified global emulator, based on inputs and decisions

A Reified influence diagram



And link, through true future values y_p, to the overall utility cost C of making decision choice d^* [Attach more models to diagram at $F^*(x^*)$]

Concluding comments

To assess our uncertainty about complex systems, it is enormously helpful to have an overall (Bayesian) framework to unify all of the sources of uncertainty.

Within this framework, all of the scientific, technical, computational, statistical and foundational issues can be addressed in principle.

Such analysis poses serious challenges, but they are no harder than all of the other modelling, computational and observational challenges involved with studying complex systems.

In particular,
Bayesian multivariate, multi-level, multi-model emulation,
careful structural discrepancy modelling
and iterative history matching
gives a great first pass treatment for most large modelling problems.

References

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: a tutorial. Reliability Engineering and System Safety 91

Kennedy, M.C. and O'Hagan, A. (2001). Bayesian calibration of computer models (with discussion). Journal of the Royal Statistical Society, B,63, 425-464

P.S. Craig, M. Goldstein, A.H. Seheult, J.A. Smith (1997). Pressure matching for hydocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments (with discussion), in Case Studies in Bayesian Statistics, vol. III, eds. C. Gastonis et al. 37-93. Springer-Verlag.

I. Vernon, M.Goldstein, and R. Bower (2010) Galaxy Formation: a Bayesian Uncertainty Analysis (with discussion), Bayesian Analysis, 5, 619-670

J. Cumming and M. Goldstein (2009) Small Sample Bayesian Designs for Complex High-Dimensional Models Based on Information Gained Using Fast Approximations, Technometrics, 51 377-388

M. Goldstein and J.C.Rougier (2008). Reified Bayesian modelling and inference for physical systems (with discussion), JSPI, 139, , 1221-1239

M. Goldstein and J.C.Rougier (2006) Bayes linear calibrated prediction for complex systems (2006), JASA, 101, 1132-1143